Weed Science 2010 58:387–394

Assessing Genetic Diversity of Canada Thistle (*Cirsium arvense*) in North America with Microsatellites

Tracey A. Bodo Slotta, Michael E. Foley, Shaioman Chao, Ruth A. Hufbauer, and David P. Horvath*

Invasive species such as Canada thistle pose a significant threat to ecosystems. The risk of introducing invasive species has increased with human activities, and the effects of such events have economic and aesthetic impacts. Native to Europe, Canada thistle is now established throughout temperate North America. Although there is documentation of early occurrences to North America, little is known on how it has become established in diverse habitats or how it continues to spread. We examined genetic diversity within and among nearly 1,700 Canada thistle individuals from 85 North American locations with the use of seven microsatellite markers in order to address these questions. PAUP and STRUCTURE programs were used to assess genetic diversity and relationships within and between populations. Populations exhibited greater within-population diversity (> 60%) than expected for a reported clonally reproducing species. Total diversity of sampled locations in North America (0.183) was less than previously reported for European locations (0.715), but the greater mean difference between North American populations (0.264 relative to 0.246 from England) suggests strong founder effects or restriction of gene flow influencing individual populations. Furthermore, analyses identified numerous instances where individuals from geographically distant regions clustered together, indicating long-distance translocation of propogules. However, isolation by distance analysis showed significant correlation between location and population genetic distances ($r = 0.1917$, $P = 0.006$). Within populations, nearly 92% of individuals sampled harbored unique multilocus genotypes, strongly suggesting that sexual reproduction is common. Within populations, analysis of genetic structure indicated significant admixture of genotypes throughout the invasive range in North America. The recurrent distribution of seed throughout North America has led to a highly diverse gene pool and increased the adaptive success Canada thistle to a wide variety of habitats. Future technologies developed for control of Canada thistle should consider this diversity.
Nomenclature: Canada thistle, *Cirsium arvense* (L.) Scop. CIRAR.
Key words: *Cirsium arvense*, Canada thistle, genetic diversity, invasive species, microsatellites.

Canada thistle has a worldwide distribution and is considered weedy in more than 25 countries (Holm et al. 1977). It is one of several invasive members of *Cirsium* that include bull thistle (*Cirsium vulagre*) and marsh thistle (*Cirsium palustre*), all with European–Mediterranean origins (Kelch and Baldwin 2003). Canada thistle was introduced to North America late in the 17th century and has since infested commercial, agricultural, residential, and natural areas (Moore 1975). It is thought that multiple and recurrent introductions of this weed have occurred (Donald 1994; McClay 2002).

The genus *Cirsium* (tribe Cardueae, family Asteraceae) occurs worldwide with approximately 80 species native to North America. Canada thistle can be distinguished from other species of *Cirsium* by the mostly dioecious flowers, horizontal root system, and tendency to grow in circular patches (Donald 1994; Hodgson 1968). Canada thistle propagates successfully both asexually through adventitious root buds and sexually with up to 5,000 seeds per plant, which may have long-term implications for maintaining diversity within and between the populations. However, previous studies suggest that asexual reproduction is important for population expansion in Canada thistle (Heimann and Cussans 1996).

The deep and extensive root system gives Canada thistle a competitive advantage over annual crops in extracting water from greater depths and it is a strong competitor for space (Donald 1994). Allelopathy may also limit competition for habitat (Bossard et al. 2000). The seeds do not require vernalization, and thus plants may emerge and flower within one growing season in warmer climates. A single seedling may spread up to 6 m in diameter in one growing season through root and secondary shoot development.

Canada thistle ranks among the top 20 invasive weeds in North America, and causes economic losses due to habitat destruction, competition with native and agricultural plants, and allelopathic effects (Donald 1994; McClay 2002). In North America, Canada thistle occurs in all states of the United States and provinces of Canada and is considered a noxious weed throughout North America, with the exception of Mexico and the southern United States (McClay 2002; U.S. Department of Agriculture, Natural Resources Conservation Service [USDA NRCS] 2005). The highest density of Canada thistle in the United States occurs in the northern Great Plains, Great Lakes, and the northeastern regions.

Canada thistle control measures have included various biological control agents (Bruckart et al. 1996; Guske et al. 2004). The efficacy and success of this approach is questionable, and some released biological control agents have had nontarget effects, harming thistles native to North America (Arnett and Louda 2002; Rand and Louda 2004). The nontarget effects of many proposed biological control agents could be due to the broad host range of introduced organisms. Furthermore, genetic differences between the introduced Canada thistle and endemic North American relatives are limited (Kelch and Baldwin 2003). Contributing to eradication difficulties, invasive Canada thistle populations may have sufficient genetic diversity or populations could contain genes conferring resistance to introduced biological control agents. Multiple or recurrent introductions of Canada thistle from Europe could further complicate the use of biological control agents for this weed. Thus, the usefulness of current biological control agents is limited.

DOI: 10.1614/WS-D-09-00070.1
* First, second, third, and fifth authors: Research Molecular Geneticist, Research Leader, Research Molecular Geneticist, and Research Plant Physiologist, U.S. Department of Agriculture, Agriculture Research Service, Red River Valley Agriculture Research Center, Fargo, ND 58102; fourth author: Associate Professor, Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO 80523-1177. Corresponding author's E-mail: tracey.slotta@gmail.com

Determinations of genetic variation, population biology, and demography for Canada thistle populations worldwide are incomplete. However, two studies have evaluated population genetic variation in the native range (Jump et al. 2003; Solé et al. 2004). Genetic diversity of established and recently founded sites of Canada thistle in Germany was evaluated with the use of amplified fragment length polymorphisms (Solé et al. 2004). They determined that site age did not contribute to diversity or even distribution of identified clones within sites and all sites in their < 5-km transect were multiclonal. With the use of 93 variable loci, they found that the proportion of samples harboring a unique genotype (and thus likely representing an individual initiated by seed rather than asexually) varied between 0.25 and 1, with a mean of 0.73. This high proportion suggests sexual reproduction plays an important role in population expansion within a site. Jump et al. (2003) employed microsatellite markers to evaluate population genetic variation in Canada thistle in the United Kingdom, with a focus on comparing central and peripheral sites. With the use of four loci, they determined that the proportion of unique genotypes ranged between 0.08 and 0.72, with a mean of 0.40. Thus, even with relatively few loci, sexual reproduction is evident, though at two sites only two genotypes were found, suggesting that those populations expanded mainly via asexual reproduction. Additionally, Jump et al. (2003) found significant isolation by distance across their UK samples.

Molecular genetic surveys similar to those in Europe have not been conducted on North American populations (Slotta et al. 2006). The invasive nature, bimodal reproductive strategy, faltering attempts at biological control, and continued success of this noxious weed across North America makes Canada thistle a focus for broad-scale sampling and analysis of populations to characterize genetic diversity, the frequency of clones in populations as indicated by shared multilocus genotypes, and geographic structuring among North American collections.

Thus, the objectives of this study were (1) to evaluate the proportion of unique genotypes within sites, as an indication of the role of sexual reproduction in population expansion within an infestation; (2) to examine genetic diversity of Canada thistle in North America; and (3) to examine genetic population structure and the strength of geographic differentiation among sites and regions. Data collected on seven variable microsatellite loci were used to address these objectives.

Materials and Methods

Plant Material Collection and DNA Extraction. Canada thistle was collected from a wide variety of habitats across North America (Figure 1 and Supplemental Table 1) for a total of 2,229 samples across 93 sites. The ratio of male to female plants within stands was estimated visually for several locations sampled by the authors. Sample habitat (e.g., rangeland and agriculture, wetland, etc.) and geographic groups were defined in Supplemental Table 1. Within sample sites, 2 to 96 individuals (mode of 15) were collected with efforts taken to ensure broad sampling of diversity by spacing collections at least 2 to 4 m apart. This intensity of sampling is roughly comparable to the sampling done by both Jump et al. 2003 and Solé et al. 2004 (though in the latter paper, small sites were sampled much more intensively). Two plots (Mplwd1 and 2) collected in Maplewood State Park (Pelican Rapids, MN) consisted of collections taken every 46 to 91 cm along six cardinal plains in plots measuring 36.6 by 36.6 m. This intensive sample was intended to examine in more detail the frequency of ramets within a naturalized location. Locations collected in Alaska were obtained from populations believed to be of recent introduction (Weeds of Alaska Database 2005).

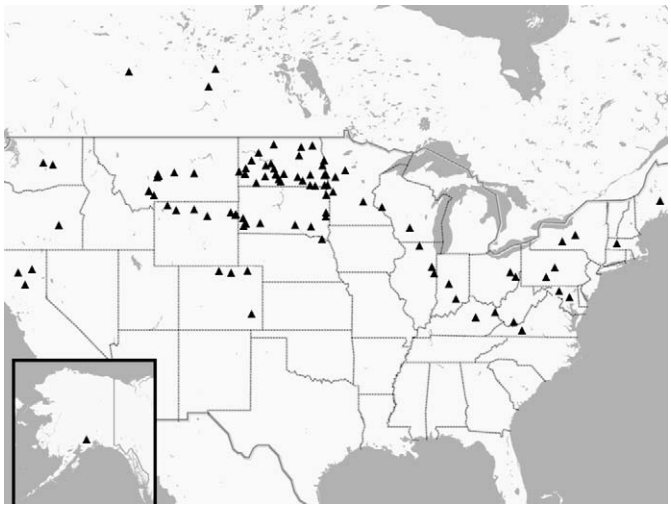

Figure 1. Distribution of sampled locations across North America.

Plant material collected was either stored at 4 to 10 C until placed at −80 C for long-term storage (2004 collections), or dried in silica gel to preserve tissue upon field collection. DNA from the 2004 collections was extracted using the DNeasy Kit[1] and from the 2005 collections with the use of a high-throughput technique (Slotta et al. 2008). DNA from samples was quantified using a spectrophotometer and aliquots diluted to 10 to 25 ng prior to amplification to ensure consistent DNA quantity among samples. Samples that did not amplify for two or more of the microsatellite loci described below or populations with fewer than seven individuals were omitted from the analysis, resulting in a total of 1,666 individuals representing 85 populations. Samples from the locations in Canada were treated as one unit.

Molecular Markers/PCR/Electrophoresis. We chose microsatellite markers, as they generally provide better fine-scale resolution for population genetic data questions than less variable markers such as cpDNA sequences (Jasieniuk and Maxwell 2001; Slotta 2008). Microsatellite regions for Canada thistle were developed as described by Slotta et al. (2005). After restriction digest products were screened by size and used to create an enriched library of the microsatellite motifs. Single recombinant cloned fragments were isolated and sequenced, resulting in 31 primer pairs from 100 screened clones. Primers were selected and tested for six individuals of Canada thistle from sample locations across North America. Of these, 28 successfully amplified and 13 were consistent in reaction product generation. With additional testing of samples, four microsatellite regions were selected (c101, c120, c128, and d117) because of the level of polymorphism among several test sampling locations. Forward primers for the four microsatellites were fluorescently end labeled with a

different dye to allow multiplexed resolution of all alleles simultaneously. An additional five primer pairs from Jump et al. (2002) were tested with reproducible polymorphic alleles for three markers (Caca 01, 05, and 10). Likewise, different fluorescent primers were prepared for these markers to facilitate simultaneous analysis. GeneScan 500 LIZ Size Standard[2] was added to each sample for estimating the size of amplified fragments. The combined products and size standard were analyzed by capillary electrophoresis with the use of an automated sequencer (ABI 3130 XL) and alleles scored with the use of GeneMapper 3.0 with a minimum peak height of 25 relative fluorescent units (rfu).

The PCR reactions contained, Buffer E[3] (Epicentre Technologies), 5 pmol each of forward-labeled and reverse primer, 10 to 25 ng of template, and 0.5 unit (U) *Taq* polymerase in 10-µl reactions (Slotta et al. 2006). Amplification conditions were 35 cycles (95 C for 30 s, 60 C for 30 s and 72 C for 45 s) followed by a final extension of 10 min at 72 C (Slotta et al. 2005) for c101, c120, c128, and d117. The amplification of the Caca series (Jump et al. 2002) followed the same conditions, except the annealing temperature was decreased to 55 C.

Analysis. A data matrix was compiled for all samples that successfully amplified for three or more of the seven microsatellite loci—a total of 1,666 individuals represented. The sizes of the amplified fragments were determined in GeneMapper and the data were visually verified prior to analysis. We calculated the proportion of unique genotypes within sample sites to evaluate whether sexual reproduction appears to play a key role in North America as observed by Jump et al. (2003) and Solé et al. (2004) for the native range.

Several analyses were done to examine genetic diversity of Canada thistle in North America. To evaluate genetic diversity of samples from different sites, the percentage of polymorphic loci, mean pairwise genetic differences (F_{ST}), and gene diversity (Nei 1987) were calculated in Arlequin (Excoffier and Lischer 2010; Schneider et al. 2000). Deviations from Hardy-Weinberg were calculated in PopGene (Guo and Thompson 1992; Yeh et al. 1999). Three tests of neutrality, which assume an infinite allele model (Table 1, Slatkin P, Ewens-Watterson FP and Chakraborty's test of neutrality P) were implemented in Arlequin and used to examine if populations were under selective pressure or if heterogosity was greater than if locations were homogenous neutral (Chakraborty 1990; Slatkin 1994; Watterson 1978). The Harpending index was used to examine demographic expansion, whether or not sampled locations were actively increasing in size based on the mismatch of random and observed haplotype differences in populations (Harpending 1994; Rogers and Harpending 1992). The Harpending index (R) predicts that as R approaches 1, the collection locations (or stands) are more likely to be expanding in size (Donnelly et al. 2001; Slatkin 1994).

Genetic population structure was examined in several ways. Evaluating how individuals from different sample sizes categorize into genetically similar clusters was determined in STRUCTURE version 2.3 (Pritchard et al. 2000). The burn-in period was set to 50,000; sample location information was used to test cluster formation. Three runs of 50,000 steps were performed for each K from 1 to 30 with allele frequencies assumed to be correlated. We evaluated consistency of these results with three additional runs with allele frequencies independent (for a total of six runs of each K). As LnP(D) leveled off well before $K = 30$ (results not shown) higher values of K were not run. To determine the most likely number of genetic clusters, we estimated the posterior probabilities of K following the STRUCTURE manual and Pritchard et al. (2001). The K with the largest posterior probability was chosen to describe the data, though lower values of K were considered as well. To evaluate more broadly variation within and among sample sites, a hierarchical analysis of molecular variance (AMOVA) was run in Arlequin. Samples were partitioned into the collected locations and the locations into geographic regions, as indicated in Supplemental Table 1. To further visualize relationships between sampled locations, cluster analysis was calculated in PAUP 4.0* (Swofford 2001) with the use of the unweighted pair-group method with arithmetical averages (UPGMA) and pairwise genetic differences based on sampled locations calculated in Arlequin (Figure 2). The length of the branches in the tree reflects genetic differences between locations.

To evaluate the strength of geographic differentiation among sites and regions, pairwise genetic distances of locations (F_{ST}) were calculated and used in examining isolation by distance. When geographically close sites are more similar to each other than to geographically distant sites, a pattern of isolation by distance emerges. This pattern arises when gene flow between adjacent sites is more common than long-distance gene flow, and when dispersal across a range occurs in a somewhat linear manner. In contrast, if there is little or no isolation by distance, then gene flow occurs over much larger distances, suggesting much larger dispersal distances. Isolation by Distance Web Service (IBDWS) was used to perform a Mantel test comparing genetic differences (F_{ST}) of sampled locations to geographic distances with the use of GPS coordinates (Jensen et al. 2005). Geographic distances between sampled locations were calculated in Mathematica with the use of a spheroidal model. Default parameters of IBDWS were used in examination of the correlation between matrices.

Results and Discussion

Genotypic Variation within Sample Locations. The seven microsatellites used in the analysis generated 82 loci for the 1,666 individuals included, representing 85 locations. The proportion of unique multilocus genotypes (e.g., plants that represent different clones) within given sample locations varied from 1 (i.e., every genotype was unique) to 0.25, with a mean of 0.92 (see Table 1). The number of polymorphic fragments varied (0 to 14) at each location, contributing to overall high genetic diversity within locations (75.7%) across North America (Table 2). Within regions, high genetic admixture is evident based on the STRUCTURE analysis (Supplemental Figure 1) with multiple multilocus genotypes present among populations from nearby localities (e.g., North Dakota [ND] and Minnesota [MN] locations and between Kentucky [KY] locations).

Reports suggest that asexual reproduction is important for population expansion in Canada thistle (Heimann and Cussans 1996). Thus, the low level of shared genotypes across North America and the high within-location diversity observed was not expected. However, several studies demon-

Table 1. Summary statistics for microsatellite variation in sampled locations.[a]

Sample location name	Number of individuals sampled	Proportion of unique genotypes	Percent polymorphic loci	Mean number pairwise differences	Gene diversity	Harpending index (*R*)	Ewens-Watterson F P value	Slatkin's exact P value	Chakraborty's test of neutrality P (*k* or more alleles)
AK1	13	0.615	10.0	0.270769	0.8246	na	0.433	0.098	0.63149
AK3	12	0.583	60.0	2.086957	0.7391	0.054	0.685	0.306	0.36275
AK4	20	1.000	100.0	4.711538	0.9603	na	0.287	0.108	0.73263
AK5	12	1.000	87.5	3.652174	0.9601	na	0	0	0.96903
AK6	16	0.250	41.7	0.814516	0.4274	0.515	0.774	0.592	0.28562
CA1	7	1.000	50.0	0.263736	0.7473	na	0.157	0.102	0.82559
CA2	15	0.571	100.0	0.43908	0.9402	na	0.059	0	0.88013
Canada	14	0.933	100.0	3.640212	0.9524	na	0	0	0.97677
CO1	82	0.929	100.0	5.305626	0.9939	na	0	0	1
CO2	7	1.012	75.0	5.67033	0.9231	na	0	0	0.94702
CO3	7	1.000	60.0	3.208791	0.8791	na	0.02	0.004	0.89634
CO4	7	0.857	92.9	6.989011	0.9231	na	0	0	0.94702
CO5	18	1.000	80.0	4.526984	0.9587	na	0.001	0	0.98508
COpc	7	1.000	75.0	4.483516	0.9231	na	0	0	0.94702
IL1	12	0.889	100.0	3.362319	0.942	na	0	0	0.96458
IL2	16	0.923	100.0	0.935484	0.8871	na	0.465	0.019	0.59237
IN1	33	0.917	83.3	2.616317	0.9828	na	0	0	0.99949
IN2	17	0.750	50.0	0.370766	0.9626	na	0	0	0.98754
KY1	18	0.970	100.0	0.590476	0.9476	0.206	0.021	0	0.92334
KY2	11	0.941	50.0	2.47619	0.6753	0.642	0.56	0.307	0.54892
KY3	18	1.000	83.3	2.4	0.946	na	0.005	0	0.95548
MA1	15	0.862	87.5	2.63908	0.9655	na	0	0	0.99034
MA2	15	0.933	75.0	1.747126	0.9563	na	0	0	0.98115
MD1	19	0.889	100.0	1.365576	0.9829	na	0	0	0.9257
MD2	14	0.455	62.5	2.455026	0.963	na	0	0	0.98812
ME	26	0.833	75.0	2.051282	0.9713	na	0	0	0.99009
MN1	30	0.933	91.7	4.332203	0.9785	na	0	0	0.99853
MN2	37	0.757	100.0	6.330989	0.9626	na	0.026	0	0.95594
MN3	12	0.833	100.0	5.56217	0.913	na	0.053	0.003	0.88113
MN4	18	0.889	100.0	3.466667	0.9524	na	0.001	0	0.95795
MN5	23	1.000	100.0	5.52657	0.9778	na	0	0	0.99808
MN6	7	1.000	70.0	3.120879	0.9231	na	0	0	0.94702
MNMPWLD	140	0.943	100.0	2.209217	0.9555	na	0	0	1
MT2	11	1.000	83.3	2.285714	0.9524	na	0	0	0.97771
MT3	6	1.000	50.0	3.030303	0.9091	na	0.001	0.001	0.93389
MT4	22	0.864	100.0	1.496829	0.9641	0.099	0	0	0.99117
ND1	9	0.780	83.3	4.96732	0.9412	na	0	0	0.96577
ND10	41	0.927	100.0	8.391147	0.9699	na	0.294	0.075	0.73599
ND11	13	1.000	100.0	8.76	0.9662	na	0	0	0.95983
ND12	18	1.111	100.0	9.498413	0.9746	na	0	0	0.98535
ND13	23	1.130	100.0	9.222222	0.9807	na	0	0	0.99141
ND14	14	1.071	100.0	8.470899	0.9577	na	0.006	0.001	0.93196
ND15	24	0.958	100.0	8.462766	0.9663	na	0.006	0	0.9542
ND16	9	1.000	83.3	6.339869	0.9542	na	0.002	0.002	0.90922
ND17	22	1.000	100.0	6.774841	0.9598	na	0.111	0.037	0.84397
ND18	10	0.900	92.9	7.410526	0.9263	na	0.001	0	0.94566
ND19	13	1.000	100.0	7.569231	0.9569	na	0.052	0.018	0.87128
ND2	18	1.000	100.0	3.625397	0.9667	na	0	0	0.98211
ND20	18	1.000	100.0	7.606349	0.973	na	0	0	0.99101
ND21	25	1.000	100.0	6.846531	0.9788	na	0	0	0.98827
ND22	20	0.950	100.0	7.265385	0.9654	na	0	0	0.9819
ND23	22	1.000	100.0	5.282241	0.9789	na	0	0	0.99339
ND24	21	1.000	100.0	4.831591	0.9338	0.015	0.019	0	0.93288
ND25	6	0.929	66.7	3.787879	0.9242	na	0.019	0.019	0.88758
ND26	24	1.000	100.0	4.705674	0.9787	na	0	0	0.99843
ND28	21	1.000	100.0	4.803717	0.971	na	0	0	0.99448
ND29	41	1.000	100.0	5.059621	0.963	na	0.1	0.001	0.87547
ND3	13	1.000	100.0	4.246154	0.9508	na	0.004	0.002	0.94896
ND4	7	1.000	66.7	4.087912	0.8901	na	0.147	0.1	0.8078
ND5	15	1.000	100.0	7.542529	0.9701	na	0	0	0.97321
ND6	13	1.000	78.6	6.578462	0.9662	na	0	0	0.95983
ND7	9	1.000	100.0	6.732026	0.9346	na	0.406	0.333	0.70979
ND8	8	1.000	90.0	4.766667	0.9417	na	0.004	0.004	0.92731
ND9	13	0.714	100.0	6.756923	0.9631	na	0	0	0.97488
NDTRNP	85	0.952	100.0	5.077341	0.9919	na	0	0	1
NY1	29	0.955	90.0	3.00605	0.9685	na	0	0	0.98592
NY2	15	0.767	100.0	2.050575	0.9563	na	0	0	0.98115
OH1	22	1.000	33.3	1.154334	0.9725	na	0	0	0.99549
OH2	30	0.933	100.0	1.868927	0.9582	na	0.01	0	0.96429
Oregon	12	0.778	100.0	0.811594	0.9529	na	0.041	0.041	0.84229
PA1	20	0.737	80.0	4.373077	0.9897	na	0.086	0.086	0.81514

Table 1. Continued.

Sample location name	Number of individuals sampled	Proportion of unique genotypes	Percent polymorphic loci	Mean number pairwise differences	Gene diversity	Harpending index (*R*)	Ewens-Watterson F P value	Slatkin's exact P value	Chakraborty's test of neutrality P (*k* or more alleles)
PA2	18	0.875	100.0	3.5	0.9762	na	0.447	0.411	0.69534
SD1	9	1.000	12.5	0.20915	0.8889	na	0.022	0.003	0.91733
SD2	19	1.000	75.0	2.475107	0.9275	na	0.049	0.002	0.89947
SD3	8	1.000	100.0	2.5	0.9333	na	0	0	0.95746
SD6	8	1.000	0.0	0	0.9333	na	0	0	0.95746
SD7	16	0.750	0.0	0	0.9677	na	0	0	0.99214
VA	12	1.000	80.0	3.913043	0.8696	0.118	0.328	0.03	0.70529
WA	14	0.826	83.3	2.37037	0.9524	0.251	0.052	0.044	0.87873
WI1	6	0.789	87.5	4.363636	0.9091	na	0	0	0.93389
WI2	23	1.000	100.0	6.025121	0.9585	na	0	0	0.97982
WI3	19	1.000	80.0	2.87909	0.9445	na	0.006	0	0.96298
WV	10	1.000	100.0	1.2	0.9474	na	0	0	0.9724
WY1	23	0.957	75.0	1.974879	0.9739	na	0	0	0.99631
WY2	21	1.000	83.3	1.728223	0.9756	na	0	0	0.99714
Mean	19.6	0.919092023	83.94257703	3.980085082	0.936177647	0.2375	0.067764706	0.032341176	0.909729529
Mode	18	1	100	0	0.9524		0	0	0.94702

[a] na, not applicable.

strated a similar level of genotypic variation in both old and recently established populations of Canada thistle (Hettwer and Gerowitt 2004; Solé et al. 2004). It was interesting to note that of the Alaska (AK) locations examined, one had the fewest number of haplotypes per individuals tested, and others were among the highest (Table 1). The least clonal Alaska location occurred near the edge of an airport. This location might be expected to have had multiple and recurrent introductions of seeds from distant locations. The sampled location with the least variation (i.e., most clonal) AK6 was collected from a landscaped setting and represented a unique genetic cluster in STRUCTURE and an isolated branch in the UPGMA tree. This population potentially represents a novel introduction to Alaska. In addition to the three Alaska collections, the other highly clonal locations include Maine (ME), ND29, and KY. The least clonal sampled locations were from small research plots in Pennsylvania (PA) and Maryland (MD) (1.05 haplotypes/individual), both used for studying the effects of biological control candidates, with an additional 14 populations indicating multiple clones per population (≥ 1.0 haplotypes/individuals in population ratio).

A more direct assessment of clonal propagation was carried out on two sampled locations collected from Maplewood State Park in MN. We found only seven shared genotypes out of the 140 plants sampled in the plots; the majority of individuals were unique genotypes. The limited number of plants with shared genotypes was often significantly far apart within the site (up to 50 m), suggesting that even they might not be exclusively clonally derived. Additional markers, such as amplified fragment length polymorphisms (AFLPs), would be needed to evaluate more conclusively whether in these cases shared alleles at our seven microsatellite loci indicated close relationships or clonality. Overall, there was limited evidence supporting clonal reproduction as the dominant mechanism for reproduction in this intensively sampled population. Diversity of the seeds that established a population, or somatic variation generated during clonal propagation could account for diversity in clonal stands. We assessed the genetic variability generated by somatic mutation in Canada thistle in 1- to 2-yr-old potted plants with multiple ramets. No somatic variation was detected among samples (car1 [a to f], car2 [b to f], car3 [a to e]), as one genetic profile was recorded for each pot (data not shown). These results indicated our methods could clearly identify clones in controlled pots as well as wild stands, with no unexpected increase in genetic diversity within ramets.

The dioecious nature of sexual reproduction of Canada thistle likely contributes to the within-stand genetic diversity. However, the ability to reproduce clonally allows stands to contain only male or female plants. We observed no single-sex stands (Supplemental Table 1) and the ratio of male to female plants within stands did not correlate with the number of individuals present based on visual surveys (Supplemental Table 1). No demographic differences were detected between the mostly male stands and the mostly female stands. These observations are consistent with the hypothesis that purely clonally propagated stands are rare. The lack of all-male or all-female stands also suggests that sexual reproduction is involved in stand maintenance.

The habitats and altitudes at which Canada thistle was collected reflect the adaptability and success in establishment under various conditions. The plants grow successfully in standing water, moist fields, roadsides, or parks, or in arid soils in a wide range of altitudes. In examining deviation from Hardy-Weinberg expectations in North America, no directed selection was detected ($P = 0.368$ based on Markov simulation of haplotype frequency) (Guo and Thompson 1992). Within sampled locations, neutrality of Slatkin and Ewens-Watterson (Table 1) indicates 33 to 46% of locations have no significant deviation from neutral selection ($P < 0.025$ or $P > 0.975$).

During periods of range expansion, weedy plants may evolve rapidly in adapting to new habitats and increase the likelihood of resistant or tolerant phenotypes (Müller-Schärer 2004), thus making effective control measures more difficult. Harpending index values (Table 1) were calculated to indicate whether stands were expanding or stable (Harpending 1994). Expansion was not computable for many locations examined because they did not have observed pairwise mismatch values greater than the expected mismatch (Table 1). However, for locations which Harpending index was calculated, most were not expanding ($R < 0.2$) and two populations indicating low to moderate range expansion ($0.5 < R < 0.75$) (KY2 and AK6).

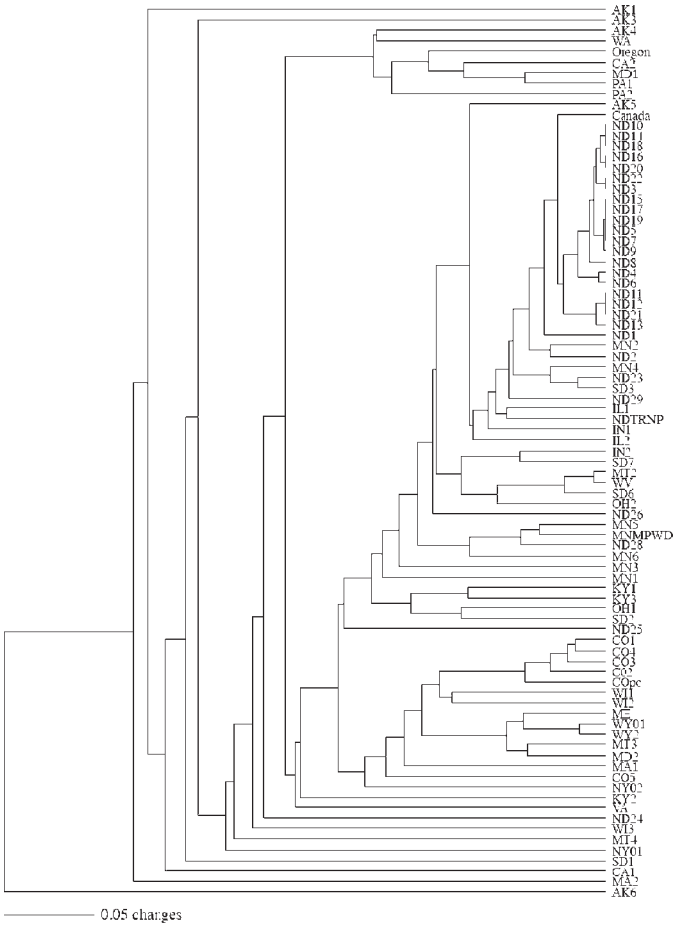

Figure 2. Cluster analysis (UPGMA) of North American Canadian thistle locations included in the study. Length of branches representative of mean pairwise genetic changes (F_{ST}) between. Locations as indicated in Supplemental file 1, all locations included separately, except Canada (three locations grouped).

Based on the level of genetic variation recovered within locations, development of technologies to control Canada thistle should focus on minimizing or eliminating seed production and not exclusively on limiting asexual reproduction. The high within-location variation discovered across diverse habitats indicates there is a need to test the effectiveness of biological control agents and chemical control options on a wide diversity of genotypes.

Variation among North American Locations. Evidence indicates considerable genetic differentiation within and between Canada thistle locations sampled throughout the United States (Tables 1 through 3). Common genotypes were detected among geographically distant locations. Multiple introductions of seeds from European or Asian locations or long-distance transport of seeds within North America may be responsible for observations that some geographically distant locations appear to share genotypes. Our data are consistent with the hypothesis that there is significant gene flow between stands in North American Canada thistle. Results from the Isolation by Distance analysis indicate the collected locations were genetically and geographically exclusive ($r = 0.1917$, $P = 0.006$). However, high genetic admixture within regions of sampled locations (e.g., KY and ND/MN in Supplemental Figure 1) is evidence of local gene flow among populations. These findings indicate that pollen flow or dispersal by seed is greater than expected. To control Canada thistle effectively, prevention of flowering, seed set, or seed dispersal may be critical.

Table 2. Diversity in Canadian thistle revealed by microsatellites. Comparison of diversity calculated based on four microsatellites in Jump et al. (2003) (1) and seven microsatellites, three of which overlap with those examined for diversity in North Dakota (Slotta et al. 2006) (2) and the current findings across North America.

	England (1)	North Dakota (2)	North America
Total diversity (H_t)	0.715	0.160	0.183
Within populations (H_s)	0.539	0.642	0.757
Among populations (G_{st})	0.246	0.026	0.264
Genetic distance between populations (mean)	0.553	0.100	0.792

A continuous distribution of Canada thistle between geographic locations is evident. Definitive boundaries between geographically isolated samples were not recovered with clusters forming from distant sampled locations (Figure 2) or with unique regional genetic clusters (Supplemental Figure 1). Among the 85 locations included in the STRUCTURE analysis, the highest posterior probability, effectively equal to 1, was obtained when $K = 18$, indicating a model of 18 genetic clusters best describes the data (Supplemental Figure 1). $K = 18$ appears to describe the data well (Supplemental Figure 1). Regions such as Indiana (IN) that did not assign clearly to a single cluster, also appeared admixed with lower values of *K*. Alpha values (admixture levels) were relatively low on average (~ 0.2) and steady across higher values of *K*, also suggesting that the analysis indicates real genetic structure (Pritchard et al. 2000, STRUCTURE manual). Within a genetic cluster, populations representing eastern North America (New York1 [NY1], PA1, ME, MD1) were genetically similar to western North American populations (Montana4 [MT4] and ND9), suggesting one area founded the other, or that there is ongoing gene flow between them. Continued introductions may also be inferred with the representation of unique genetic clusters as found with AK6, which also is isolated outside the remaining samples in the UPGMA analysis. The remaining samples from Alaska did not group together in UPGMA clustering (Figure 2) or represent a single genetic group in the STRUCTURE analysis (Supplemental Figure 1). Likewise, sampled locations from the east coast clustered with those from the west coast, for example MD2 and ME cluster with Wyoming01 (WY01), WY2 and MT3 in the UPGMA, which evaluates data at the population level rather than the individual level used by STRUCTURE. Again, this indicates that there may be ongoing gene flow between these areas, perhaps via transport of agricultural or other products between the coasts. Given that AK4 was collected adjacent to the airport in Anchorage, AK, might suggest the latter. Sources for the recently introduced AK plants can be attributed to the Midwest (Illinois/Indiana), southeast (Virginia/Kentucky), and north-

Table 3. Analysis of molecular variance for North American populations.

Level	Diversity	Variance
Within individuals	8.33277	$F_{it} = 0.91667$*
Within populations	60.96606%	$F_{is} = 0.87976$*
Among populations within groups	22.40424%	$F_{sc} = 0.24431$*
Among groups	8.29693%	$F_{ct} = 0.08297$*

* $P = 0$.

west (Washington), with the Alaskan populations being grouped with Midwestern populations in the STRUCTURE analysis. Taken together, these data suggest that distribution of pollen or seeds among locations in North America could be problematic and that evolution of resistance to chemical or biological control agents could readily spread throughout the country and perhaps the world.

Total diversity of Canada thistle was calculated for North American samples based on seven microsatellite loci; three of these loci were used, in addition with one other loci, in estimating diversity of Canada thistle plants in England (Jump et al. 2003) (Table 2). Lower diversity was recovered in the current analysis of North American plants than those sampled in England. This may result from European populations being closer to the center of origin, or an artifact of different markers exhibiting different levels of variation. However, multiple sources of genetic material from all over the world, greater land mass and habitat variation, and large population sizes may contribute to the higher level of within-population diversity observed in North America. This diversity will make Canada thistle control more difficult, as novel genotypes in new habitats may limit control options (Parker et al. 2006).

Continued human activities, such as incidental seed transportation, are leading to the establishment of additional individuals and increasing the genetic diversity in existing populations. The number of individuals in initial introductions to new locations has provided the baseline genetic diversity, and repeated introductions greatly increase the probability of adaptive evolution of weedy species (Gaskin et al. 2005; Marrs et al. 2008). Indeed, given the possibility of multiple introductions from populations around the world in Canada thistle, multiple locations in Asia and Europe should be included in the search for biological control agents to provide greater diversity of options. Biological control agents from the same initial range as founder populations are the most logical choice to explore in developing new technologies. Invertebrate herbivores such as gall flies (*Uphora* spp.) and weevils (*Rhinocyllis* spp.) or pathogens (e.g., *Puccinia* or *Pseudomonas syrningae* strains) that specialize in attacking Canada thistle or other European thistles may be the most suitable option to controlling this ever-increasing problem in North America (Parker et al. 2006).

Beyond Canada thistle, the implications of these findings to invasive-species biology indicate more knowledge is needed on basic biology of weedy species in developing effective control measures. Multiple introductions and continued global transportation in weedy species leads to maintenance of genetic diversity and adaptive success, thereby increasing the difficulty in developing effective management strategies (Gaskin et al. 2005; Marrs et al. 2008; Müller-Schärer et al. 2004). Populations or locations may not only differ in habitat characteristics, but in genetic variation, expansion capability, and reproductive success. With the greatest degree of variation housed within populations, new biological control or chemical control options for weedy invasives need to consider the adaptability of plants to the intense selective pressures applied during eradication (Lambrinos 2004).

Sources of Materials

[1] DNeasy Mini Kit, Qiagen Inc., 27220 Turnberry Lane, Suite 200, Valencia, CA 91355.

[2] GeneScan 500 LIZ Size Standard, Applied Biosystems Inc., 850 Lincoln Centre Drive, Foster City, CA 94404.

[3] Buffer E, EPICENTRE Biotechnologies, 726 Post Road, Madison, WI 53713.

Acknowledgments

Critical lab assistance was provided by Jennifer Jewett, Laura Kelley, and Jamie Rust. Sample material was obtained through cooperation with the National Parks Service and Minnesota State Parks. Collaborators who assisted in collecting samples include P. Bhomik, S. Boyetchko, B. Bruckhart, K. Christianson, B. Curran, E. Gallandt, J. Gaskin, J. Jewett, L. Kelley, J. Lydon, R. Lym, L. Samuel, R. Shelley, J. Thompson, and B. Villegas. Valuable comments on earlier drafts were provided by R. Roehrdanz and J. Gaskin.

Literature Cited

Arnett, A. E. and S. M. Louda. 2002. Re-test of *Rhinocyllus conicus* host specificity and the prediction of ecological risk in biological control. Biol. Conserv. 106:251–257.

Bossard, C. C., J. M. Randall, and M. C. Hoshovsky. 2000. Invasive Plants of California's Wildlands. Berkeley: University of California Press.

Bruckart, W. L., D. J. Poltis, G. DeFago, S. S. Rosenthal, and D. M. Supkoff. 1996. Susceptibility of *Carduus*, *Cirsium* and *Cynara scolymus* species artificially inoculated with *Puccinia carduorum* from musk thistle. Biol. Control 6:215–221.

Chakraborty, R. 1990. Mitochondrial DNA polymorphism reveals hidden heterogeneity within some Asian populations. Am. J. Hum. Genet. 47:87–94.

Donald, W. W. 1994. The biology of Canada thistle (*Cirsium arvense*). Rev. Weed Sci. 6:77–101.

Donnelly, M. J., M. C. Licht, and T. Lehmann. 2001. Evidence for recent population expansion in the evolutionary history of the malarial vectors *Anopheles arabiensis* and *Anopheles gambiae*. Mol. Biol. Evol. 18:1353–1364.

Excoffier, L. and H.E.L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. In press.

Gaskin, J. F., D. Y. Zhang, and M. C. Mon. 2005. Invasion of *Lepidium draba* (Brassicaceae) in the Western United States: distributions and origins of chloroplast haplotypes. Mol. Ecol. 14:2331–2341.

Guo, S. W. and E. A. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372.

Guske, S., B. Schultz, and C. Boyle. 2004. Biocontrol options for *Cirsium arvense* with indigenous fungal pathogens. Weed Res. 44:107–116.

Harpending, H. C. 1994. Signature of ancient population growth in low-resolution mitochondrial DNA mismatch distribution. Human Biol. 66:591–600.

Heimann, B. and G. W. Cussans. 1996. The importance of seeds and sexual reproduction in the population biology of *Cirsium arvense*—a literature review. Weed Res. 36:493–503.

Hettwer, U. and B. Gerowitt. 2004. An investigation of genetic variation in *Cirsium arvense* field patches. Weed Res. 44:289–297.

Hodgson, J. M. 1968. The nature, ecology, and control of Canada thistle. Washington, DC: U.S. Department of Agriculture, Agricultural Research Service, Tech. Bull. 1386.

Holm, L. G., D. L. Plucknett, J. V. Pancho, and J. P. Herberger. 1977. The World's Worst Weeds—Distribution and Biology. Honolulu: University of Hawaii Press.

Jasieniuk, M. and B. D. Maxwell. 2001. Plant diversity: new insights from molecular biology and genomics technologies. Weed Sci. 49:257–265.

Jensen, J. L., A. J. Bohonak, and S. T. Kelley. 2005. Isolation by distance, web service. BMC Gen 6, 13. v.3.15. http://ibdws.sdsu.edu/.

Jump, A. S., D. H. Dawson, C. M. James, F. I. Woodward, and T. Burke. 2002. Isolation of polymorphic microsatellites in the stemless thistle (*Cirsium acaule*) and their utility in other *Cirsium* species. Mol. Ecol. Notes 2:589–592.

Jump, A. S., F. I. Woodward, and T. Burke. 2003. *Cirsium* species show disparity in patterns of genetic variation at their range-edge, despite similar patterns of reproduction and isolation. New Phytol. 160:359–370.

Kelch, D. G. and B. G. Baldwin. 2003. Phylogeny and ecological radiation of New World thistles (*Cirsium*, Cardueae—Compositae) based on ITS and ETS rDNA sequence data. Mol Ecol. 12:141–151.

Lambrinos, J. G. 2004. How interactions between ecology and evolution influence contemporary invasion dynamics. Ecol. 85:2061–2070.
Marrs, R. A., R. Sforza, and R. A. Hufbauer. 2008. Evidence for multiple introductions of *Centaurea stoebe micranthos* (spotted knapweed, Asteraceae) to North America. Mol. Ecol. 17:4197–4208.
McClay, A. S. 2002. Canada thistle. Page 413 *in* R. Van Driesche, B. Blossey, M. Hoddle, S. Lyon, and R. Reardon. Biological control of invasive plants in the Eastern United States. U.S. Department of Agriculture Forest Service Publication FHTET-2002-04.
Moore, R. J. 1975. The biology of Canadian weeds. 13. *Cirsium arvense* (L.) Scop. Can J Plant Sci. 55:1033–1048.
Müller-Schärer, H., U. Schaffner, and T. Seinger. 2004. Evolution in invasive plants, implications for biological control. Trends Ecol Evol. 19:417–422.
Nei, M. 1987. Molecular Evolutionary Genetics. New York, NY: Columbia University Press.
Parker, J. D., D. E. Burkepile, and M. E. Hay. 2006. Opposing effects of native and exotic herbivores on plant invasions. Science 311:1459–1461.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.
Rand, T. A. and S. M. Louda. 2004. Exotic weed invasion increases the susceptibility of native plants to attack by a biocontrol herbivore. Ecology 85:1548–1554.
Rogers, A. R. and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. 9:552–569.
Schneider, S., D. Roessli, and L. Excoffier. 2000. Arelquin Version 2.000: A Software for Population Genetics Analysis. Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva.
Slatkin, M. 1994. An exact test for neutrality based on the Ewens sampling distribution. Genet. Res. 64:71–74.
Slotta, T.A.B. 2008. What we know about weeds: insights from genetic markers. Weed Sci. 56:322–326.
Slotta, T.A.B., L. Brady, and S. Chao. 2008. High throughput tissue preparation for large-scale genotyping experiments. Mol. Ecol. Res. 8:83–87.
Slotta, T.A.B., D. P. Horvath, and M. E. Foley. 2005. Development of polymorphic markers for *Cirsium arvense*, Canada thistle, and their amplification in closely related taxa. Mol. Ecol. Notes 5:917–919.
Slotta, T.A.B., J. M. Rothhouse, D. P. Horvath, and M. E. Foley. 2006. Genetic diversity of Canada thistle (*Cirsium arvense*) in North Dakota. Weed Sci. 54:1080–1085.
Solé, M., W. Durka, S. Eber, and R. Brandl. 2004. Genotypic and genetic diversity of the common weed *Cirsium arvense* (Asteraceae). Int. J. Plant Sci. 165:437–444.
Swofford, D. L. 2001. PAUP* Phylogenetic Analysis Using Parsimony, ver. 4.0b10. Sunderland, MA: Sinauer Associates.
U.S. Department of Agriculture, Natural Resources Conservation Service. 2005. The PLANTS Database, version 3.5 (http://plants.usda.gov). Data compiled from various sources by Mark W. Skinner. Baton Rouge, LA: National Plant Data Center.
Watterson, G. 1978. The homozygosity test of neutrality. Genetics 88:405–417.
Weeds of Alaska Database. 2005. Database of exotic vegetation collected in Alaska. University of Alaska, Alaska Natural Heritage Program—US Forest Service—National Park Service Database. http://akweeds.uaa.alaska.edu/. Accessed: June 7, 2010.
Yeh, F. C., R. C. Yang, and T. Boyle. 1999. PopGene ver. 1.31, Microsoft Window-based Freeware for Population Genetic Analysis. Free program distributed by the authors. http://www.ualberta.ca/~fyeh/. Accessed: June 7, 2010.

Received November 10, 2009, and approved May 1, 2010.